VU VRIJE UNIVERSITEIT AMSTERDAM

VU Research Portal

The Center for Epidemiologic Studies Depression scale (CESD-D) in an mixed mode, repeated measures design: age and sex differences

Geerlings, S.W.; Beekman, A.T.F.; Deeg, D.J.H.; Smit, J.H.; van Tilburg, W.

published in
International Journal of Methods in Psychiatric Research
1999

DOI (link to publisher)
10.1002/mpr.60

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
Geerlings, S. W., Beekman, A. T. F., Deeg, D. J. H., Smit, J. H., & van Tilburg, W. (1999). The Center for Epidemiologic Studies Depression scale (CESD-D) in an mixed mode, repeated measures design: age and sex differences. *International Journal of Methods in Psychiatric Research*, *8*, 102-109. https://doi.org/10.1002/mpr.60

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 05. May. 2023

The Center for Epidemiologic Studies Depression scale (CES-D) in a mixed-mode repeated measurements design: sex and age effects in older adults

SANDRA W. GEERLINGS, AARTJAN T.F. BEEKMAN, DORLY J.H. DEEG, WILLEM VAN TILBURG, Department of Psychiatry, Free University Amsterdam, Netherlands
JAN H. SMIT, Department of Sociology and Social Gerontology, Free University Amsterdam, Netherlands

ABSTRACT *In order to obtain repeated measurements of depression in an efficient and relatively inexpensive design, a mixture of face-to-face interviews and mail questionnaires was employed. The aims of the study were to examine mode effects of face-to-face interviews versus mail questionnaires on depression scores and to test potential interactions between mode of data collection and sex and age of the respondents. In the study sample, which at the outset consisted of 327 depressed and 325 non-depressed older adults (55–85 years) drawn from a larger random community based sample in the Netherlands, depression was measured in successive waves (cycles), using the Center for Epidemiologic Studies of Depression scale (CES-D). With mode of data collection and sex and age of the respondents as independent variables, differences in CES-D scores were analysed. The CES-D scores were higher when collected by mail questionnaires than when face-to-face interviews were used. No systematic interactions between sex and age of the respondents with mode of data collection were found. For the scores based on mail questionnaires, a transformation is proposed, resulting in scores that are comparable to those obtained by interviews. In studying depression in older adults, more cost-effective mail questionnaires may be used in addition to face-to-face interviews, provided that a transformation is performed before embarking on the analysis.*

Key words: depression, older adults, longitudinal studies, mode of data collection, transformation.

Introduction

Depression is considered to be one of the most prevalent psychiatric disorders among older adults (Blazer, 1994). This has led to a growing number of studies in which the impact of potential risk factors on both the onset and the course of depression in older adults is evaluated. These studies aim at a better understanding of depression, which may help to design more effective strategies of prevention and treatment. In general, the most appropriate design to evaluate the impact of risk factors on the onset and the course of depression is a large-scale prospective longitudinal community-based study in which the development of depression is investigated by repeatedly measuring respondents' depression states over time. The mode commonly used for collecting data on depression is the face-to-face interview, but applying this mode repeatedly in large-scale prospective longitudinal community based studies is expensive. Budget constraints have led to designs in which data are collected with mixed modes: less expensive data collection modes such as mail questionnaires and telephone interviews are used in combination with the more expensive face-to-face interviews. However, using mixed-mode data collection procedures may reduce comparability of the data collected, as the mode may influence the respondents' answer behaviour, introducing different sources of bias while estimating depression (Groves, 1989; De Leeuw, 1992).

The present study focuses on mode effects in data on depression in a large-scale, prospective community-based study among older adults in which data were collected with mixed modes: face-to-face interviews and mail questionnaires. Systematic differences between data concerning depression derived from face-to-face interviews and mail questionnaires may be expected. Reporting depression is related to disclosure mechanisms. Mail questionnaires are answered in a

more anonymous context resulting in fewer barriers in reporting negative self-presentations and thus a greater willingness to reveal sensitive information (Cannel and Fowler, 1963; Bradburn, 1983; Sudman and Bradburn, 1974; Schwarz, Strack, Hippler and Bishop, 1991). This may result in reporting more depressive symptoms in mail questionnaires compared with face-to-face interviews.

Mail questionnaires may yield more candid reactions on sensitive questions concerning depression than face-to-face interviews. However, it is generally assumed that questions on depression are difficult to answer, and face-to-face interviews are superior to mail questionnaires in explaining procedures, solving misunderstandings and reducing (item) non-response (De Leeuw, 1992). These advantages of face-to-face interviews over mail questionnaires have led to the opinion that the former mode may be superior in terms of validity in depression research.

Studies on depression comparing face-to-face interviews with mail questionnaires have led to controversial results. In some studies, mail questionnaires yielded significantly higher depression scores (Baumgart and Oliver, 1981; Aquilino, 1998), whereas in other studies no such mode effect was found (King and Buchwald, 1982; Bryson and Pylon, 1984). It has also been suggested that there is an interaction between the mode of data collection and the sex of the respondent. In face-to-face interviews males might be less likely than females to reveal depression because of the fear of receiving a negative evaluation from the interviewer (Baumgart and Oliver, 1981). Mode effects on sensitive topics such as depression may also be related to the age of the respondents. Although empirical studies on mode effects among older adults are scarce, there are indications that especially older respondents' answer behaviour differs in mail and face-to-face data collection procedures (Herzog and Kulka, 1989).

In the present study, data on the Center for Epidemiologic Studies Depression scale (CES-D) (Radloff, 1977) were collected at eight successive waves over a period of three years. In the first and the last wave, data were collected using face-to-face interviews, and in the intermediate period mail questionnaires were used. This procedure enables the study of a mode effect. In the case of a mode effect, a transformation procedure of the answers obtained from the mail questionnaires is proposed, resulting in scores comparable to those obtained from the face-to-face interviews

Methods

Sampling and procedures

The present study is part of the Longitudinal Aging Study Amsterdam (LASA), which is a 10-year interdisciplinary study on predictors and consequences of changes in autonomy and well-being in the aging population (Deeg and Westendorp de Serière, 1994). Full details on sampling and response have been described elsewhere (Beekman, Deeg, Van Limbeek, Braam, De Vries, Van Tilburg, 1997). At baseline (t_0), 3056 interviews were available. On the CES-D (Radloff, 1977), 454 respondents (14.9%) had a score of at least 16, which is the generally accepted cut-off score for a clinically relevant depressive syndrome (Berkman, Berkman, Kasl, Freeman, Leo, Ostfield, Coroni-Huntly and Brody, 1986; Beekman et al., 1997). All depressed subjects, and a random sample of non-depressed subjects were asked to participate in a side study concentrating on late life depression, which involved a diagnostic interview at the beginning and a series of follow-up measurements by means of mail questionnaires. The sample participating in this side study consisted of a depressed cohort (n = 327) and a non-depressed cohort (n = 325). Non-response was associated with age ($p < 0.001$), but not with sex. Follow-up data on depressive symptoms were collected at five-monthly intervals in seven successive waves (t_1 to t_7), covering a period of three years. At t_1 to t_6, data on depressive symptoms were collected by mail questionnaires, and in the final wave (t_7), the same face-to-face mode of interviewing was used as at t_0. For the present study, we used CES-D data from the first and the last wave (t_0 and t_7, both interview data), and those from the second and the penultimate wave (t_1 and t_6 both mail questionnaire data). Response at t_1, at t_6 and at t_7 was 74.5%, 66.6% and 72.1%, respectively. Attrition after t_0 was associated with age ($p < 0.001$) and depression ($p < 0.001$) but not with sex.

Measurements

Depressive symptoms were measured, using the Center for Epidemiologic Studies Depression scale (CES-D) (Radloff, 1977). This is a 20-item self-report scale developed to measure depressive symptoms in the community. The 20 symptom items constitute four subscales: depressed affect, positive affect, somatic complaints, and interpersonal problems (Radloff, 1977; Smit, Van den Eeden, Deeg, Beekman, 1995). Subjects were asked how often they experienced each symptom during the previous week. Items were scored

on a four-point scale, ranging from 0 (rarely or none of the time) to 3 (most of or all the time). The values of these response categories were reversed for the positive affect items. The total CES-D score ranges from zero to 60. A score of 16 or higher is generally used to indicate the presence of a clinically relevant depressive syndrome (Berkman et al., 1986; Beekman et al., 1997). The psychometric properties of the scale were found to be good in older populations (Radloff and Teri, 1986; Beekman, Van Limbeek, Deeg, Wouters, Van Tilburg, 1994).

At t_0 and t_7 CES-D data were collected using face-to-face interviews and at t_1 to t_6 mail questionnaires were used. The face-to-face interviews were administered by trained interviewers at the respondents' residences. The respondents were requested to answer the questions, while a card listing the response categories and their codes were presented. The interviewers were trained to restrict their interference with the interview to a minimum, and the interviews were audiotaped in order to perform quality control on interviewer behaviour. In the mail questionnaires, the question order, the question wording and the response categories and their codes were identical to those in the face-to-face interviews. In both the face-to-face interviews and the mail questionnaires, the reliability indexed by Cronbach's alpha was high (≥ 0.89).

Data-analysis

In the LASA baseline sample (t_0) 14.9% of the respondents scored 16 or higher on the CES-D and in the present study sample this was stratified to be 50%. Due to this sampling procedure, inverse sampling probability weights (respondents scoring <16 : respondents scoring ≥16 = 85.1% : 14.9%) had to be applied in order to achieve a sample which is representative for the original sample. The non-depressed subjects were given a weight of 1.70 and those depressed at baseline were given a weight of 0.30. In all stages of the analysis, CES-D (scale) scores, collected by two different modes, were the dependent variables.

CES-D data were collected by different modes at successive waves with five-monthly intervals over a period of three years, and thus in studying mode effects time-related changes in CES-D scores had to be taken into account. Real changes and artefactual changes can be distinguished. In this study, only scores at two successive waves with intervals of five months were compared: CES-D scores at t_0 were compared with those at t_1 and CES-D scores at t_6 were compared with those at t_7. As it is unlikely that the average level of depression changes significantly over a period of five months in older adults in the general population, real changes were not expected between two measurements. However, artefactual changes caused by selective attrition during the study and by repeated measurement (retest effects) had to be taken into account. Furthermore, a sequence effect, due to the fact that in the first two waves an interview (t_0) was followed by a mail questionnaire (t_1) and in the last two waves the sequence was reversed (t_6 mail questionnaire; t_7 interview), had to be accommodated for. In order to cope with the above, two separate analyses were performed. In the first analysis, where potential bias due to selective attrition was avoided by using the largest possible number of respondents (n = 519), CES-D scores at t_0 were compared with those at t_1. For the second analysis, where potential bias due to retest effects and sequence effects was ruled out, we averaged CES-D scores at t_0 and t_7 (both interview data) and at t_1 and t_6 (both mail questionnaire data) for respondents with valid observations at all four occasions (n = 369).

Data were analysed, using multivariate analysis of variance (MANOVA) for repeated measurements, with mode (face-to-face interview versus mail questionnaire) as within-subjects-factor and sex and age of respondents at t_0 as between-subjects factors. Age was dichotomized at the median (≤70/>70). As the CES-D scale is highly skewed with a minimum of zero, all analyses were repeated, using a logarithmic transformation of the scale score plus one.

To examine whether a mode effect exists on each of the four subscales of the CES-D (Radloff, 1977), the same analyses were repeated at the subscale level (depressed affect, positive affect, somatic complaints, and interpersonal problems). In preliminary analysis, both in the face-to-face interview data and in the mail questionnaire data, the presence of the four subscales was examined using factor analysis (analysis of principal components with varimax rotation).

In the case of a mode effect (higher depression scores in the mail questionnaires than in the face-to-face interviews), a transformation is proposed in order to make the CES-D data collected by mail questionnaires (t_1 to t_6) comparable with the data collected by interviews (t_0 and t_7). Comparability of the scores at t_0 to t_7 after transformation is defined as similarity with respect to important test parameters such as the mean, the standard deviation and the percentage scoring ≥16 the

percentage scoring at the cut-off score of 16 or above, thereby preserving interpretable scores and not affecting the original inter-individual differences. These conditions are most likely to be fulfilled when using a generalized T-score transformation (Nunnally, 1978; Kohout, Berkman, Evans, Cornoni-Huntley, 1993). This transformation involves three steps: first, the standardizing (*z*-score) transformation is applied; next, each *z*-score is multiplied by the standard deviation of the criterion distribution (CES-D at face-to-face interview); and finally, the mean of the criterion distribution is added to each score. In the absence of significant differences between the means and standard deviations of the CES-D at t_0 and t_7, the mean and the standard deviation at t_0 are used as criterion parameters. In the presence of differences, adjustments will have to be made.

In formula:

$$X_{trans} = ((X_{ti} - \mu_{ti})/ \acute{o}_{ti}) * \acute{o}_{t0} + u_{t0}$$

where X_{trans} = scores on the transformed scale
i = wave number of mail questionnaire (1 – 6)
X_{ti} = scores on the original scale at t_i
μ_{ti}; $\acute{o}_{ti}$ = mean and standard deviation of X_{ti}
μ_{t0}; $\acute{o}_{t0}$ = mean and standard deviation of the criterion distribution at t_0

Results

In Table 1, mean CES-D scores are presented for the weighted sample. Comparison of mean CES-D scores at t_0 (7.81) with those at t_1 (9.85; difference: 2.04) and those at t_6 (10.55) with those at t_7 (7.92; difference: 2.63) suggests a mode effect. The averaged scores of t_0 and t_7 (both interview data) and of t_1 and t_6 (both mail questionnaire data) were 7.07 and 9.75, resulting in a difference of 2.68 ($t = 10.07$, $df = 368$, $p < 0.001$).

Table 1: CES-D scores at t_0–t_7 in weighted study sample

	Mode	n	mean (std dev)	median	range	%≥16
t_0	interview	652	7.81(7.82)	5.00	0–48	15.0
t_1	mail	519	9.85(7.93)	8.00	0–52	18.8
t_2	mail	541	9.49(7.99)	8.00	0–51	18.8
t_3	mail	513	10.64 (8.12)	9.00	0–47	25.3
t_4	mail	497	10.27 (8.36)	8.00	0–46	23.0
t_5	mail	400	10.22 (8.01)	9.00	0–45	23.3
t_6	mail	463	10.55 (8.33)	9.00	0–42	23.7
t_7	interview	486	7.92 (7.65)	6.00	0–46	14.8

Table 2 presents the results of the first MANOVA analysis. Instead of a full factorial model, a model without the interaction term of sex and age is given, as there was no significant interaction between these factors. The demographic variables sex and age at t_0 had an effect on CES-D scores. Women and older people had relatively high depression scores both at t_0 and at t_1. The mode effect was significant: CES-D scores were higher at t_1 than at t_0. No interactions between demographic characteristics and mode of data collection were found. In the logarithmic transformation condition, the results were essentially the same (data not shown). Results of the second MANOVA analysis are presented in Table 3. The results are similar to those of the first analysis. In the logarithmic transformation condition, the results were essentially the same as those of the first analyses (data not shown). To check whether differences in variances among subgroups affected the results, stratified analyses assessing mode effect among younger old and older old males and females were performed. In all analyses, the same (highly significant) mode effect was found (data not shown). For all subscales a significant mode effect was found, the most prominent of which applied to the positive affect subscale (Table 4).

Table 2: Mean CES-D scores at t_0 and t_1 for sex, age and mode

	Mode	
	Interview mean	Mail mean
Total	7.15	9.85
Sex Male		
Age≤70	5.38	7.14
> 70	5.75	9.45
Sex Female		
Age≤70	7.55	10.02
> 70	10.52	13.57
		$F_{1,515}$ 70.70***

Significance tests are F ratios from MANOVA: mode effect (see table); sex effect: $F_{1,515}=34.36$, $p < 0.001$); age effect: $F_{1,515}= 15$ is 89, $p < 0.001$). The interaction of mode and demographic variables (sex and age) did not reach the level of significance ($p = 0.05$).

*** $p < 0.001$

The mail questionnaire data (t_1 to t_6) were transformed by a T-score transformation. As there were no systematic interactions between mode effect and sex and age of the respondents, one transformation was performed for male and female subjects in all age groups. Thus, CES-D scores were transformed, with no adjustment made for sex and age. The mean and the standard deviation at t_0 and at t_7 were essentially the same (Table 1), so only those at t_0 were used as criterion parameters. Results of the transformation are presented in Table 5. As intended, important characteristics of the CES-D scores, such as the mean and standard deviation and the percentage scoring ≥16 have become very similar.

Table 3: Mean of averaged CES-D scores at t_0/t_7 and t_1/t_6 for sex, age and mode

	Mode	
	Interview mean	Mail mean
Total	7.07	9.75
Sex Male		
Age ≤ 70	4.57	6.95
> 70	5.51	8.42
Sex Female		
Age ≤ 70	7.91	10.99
> 70	11.38	13.52
		$F_{1,365}$ =90.63***

Significance tests are F ratios from MANOVA: mode effect (see table); sex effect: $F_{1,365}$=55.85, p< 0.001); age effect: $F_{1, 365}$ = 12.14, p < 0.01). The interaction of mode and demographic variables (sex and age) did not reach the level of significance (p = 0.05).
*** p < 0.001.

Discussion

The need for prospective longitudinal community-based studies on depression in older adults leads to the question of whether there are economical ways of data collection that may be used in addition to face-to-face interviews. Data collected by different modes were therefore compared. In the present study, mode effects of face-to-face interviews versus self-administered mail questionnaires on CES-D scores were examined. As expected, differences in CES-D scores were observed between the two data-collection modes. CES-D scores were significantly higher when they were collected by mail questionnaires than when face-to-face interviews were used. These results replicate the findings of an earlier study using the same scale to assess symptom severity in younger adults (Aquilino, 1998). The differences were demonstrated in each of the four subscales, and they were particularly notable in the positive affect items (such as 'I felt hopeful about the future'). It might be argued that positive formulations exert emotional or psychosocial pressure on the respondent to choose an answer in the direction in which the question is formulated, more than negative formulations do (Smit et al., 1995). In the context of the CES-D, and in line with our findings, this may mean that respondents in an interview setting will tend to respond relatively more positively (and thus possibly in a more socially desirable way) on the positive affect items than in mail questionnaires, where no such emotional/psychosocial pressure can be exerted. Concerning the impact of sex and age on mode effects, no systematic interactions between these characteristics and mode of data collection were found. Thus, the mode effect was the same for males and females in all age groups.

Considering the results, it is clear that if CES-D data collected by mail questionnaires and by face-to-face interviews are to be used in one study, a transformation is necessary. A mode effect was found on each of the four subscales of the CES-D, and thus the total scale was transformed. A generalized T-score transformation was shown to be satisfactory (Nunnally, 1978; Kohout et al., 1993). As mean CES-D scores at the beginning and at the end of the follow-up were essentially the same, thereby indicating stability over time, the transformation yielded equal mean depression scores at successive waves. However, the transformation did preserve variability at the individual level, which is the matter of interest in most studies. The mail questionnaire data were transformed using the face-to-face interview data as a criterion. Face-to-face interviews are superior over mail questionnaires in explaining procedures, solving misunderstandings and reducing (item) non-response (De Leeuw, 1992). These advantages probably make them more valid and therefore more useful in depression research. There is no external (independent) criterion on the basis of which one can choose between the two modes. Mail

Table 4: CES-D subscale scores for mode

	Mode			
	Interview	Mail	Interview	Mail
	t_0	t_1	t_0/t_7	t_1/t_6
Subscale	mean	mean	mean	mean
I Depressed affect	1.51	1.70	1.47	1.74
II Positive affect	2.93	4.83	2.89	4.60
III Somatic complaints	2.51	3.00	2.45	2.93
IV Interpersonal problems	0.22	0.31	0.17	0.33

Significance tests are F ratios from MANOVA with mode as within- subjects-factor.
*** p<0.001.
** p<0.01.
* p<0.05.

Table 5: CES-D scores at to-t_7 in weighted sample after transformation

	Mode	mean (stddev)	median	range	%≥16
t_0	interview	7.81(7.82)	5.00	0–48	15.0
t_1	mail	7.85(7.91)	6.00	–2–49	14.5
t_2	mail	7.64(7.82)	6.00	–1–48	14.0
t_3	mail	7.71(7.89)	6.00	–2–43	16.1
t_4	mail	7.80(7.82)	6.00	–2–41	14.3
t_5	mail	7.93(7.66)	7.00	–2–42	13.6
t_6	mail	7.83(7.75)	6.00	–2–37	15.7
t_7	interview	7.92(7.65)	6.00	0–46	14.8

questionnaires are answered in a more anonymous context, which may result in fewer barriers in revealing sensitive information than in face-to-face interviews. However, using the interview data as a criterion is more in line with clinical practice, and makes the data more comparable with other epidemiological studies in which the measurement of depression is generally incorporated in large-scale interviews. Furthermore, the CES-D was originally designed for use in surveys involving face-to-face interviews; the original norms were obtained within the context of an interview, and the data derived from interviews have generally been used in validation studies (Radloff, 1977; Beekman et al., 1994, 1997). The utility of mail questionnaires is, so far, less well understood. Finally, the proposed transformation is a conservative one (CES-D scores are decreased instead of increased), which precludes overestimations in studies on the onset and the course of depression.

Regarding the present study, some limitations must be acknowledged. First, the results may have been affected by the loss of subjects in the course of the study. It was found that the older subjects and those more depressed at the baseline measurement were more likely to be lost at follow-up. However, the adverse consequences are probably limited, as it was demonstrated that the analyses of the maximum sample lead to essentially the same conclusions as comparable analyses based on the more selective sample. Secondly, the LASA study was not specifically designed to study mode effects in older adults. To reduce confounding and to examine retest effects and sequence effects more thoroughly, an experimental design in which respondents are randomly allocated to

the questionnaire or the face-to-face interview format is recommended for future studies. In previous studies on mode differences, effects due to sampling differences have often been confounded with mode effects (Aquilino, 1998). In the present study, in which mode effects were examined within the same respondents, this potential source of bias was completely ruled out. Finally, weighted data were analysed using the SPSS statistical package, which may affect the computation of standard errors and might lead to erroneously significant findings (Heeringa and Liu, 1998). However, in this study, the adverse consequences of the use of SPSS with weighted data are probably limited. With depression data on the first and the second wave, the mode effect was also studied using Stata, a program that incorporates weighting effects in the estimation of standard errors. In the absence of the option of using multiple analysis of variance (MANOVA) for repeated measurements, a linear regression was applied. The results did not affect the outcome of this study.

In conclusion, this study adds new information to research on the measurement of depression. Until now, most studies on depression rely on data obtained by cross section, whereas some have measured depression at two successive waves. However, in order to obtain a reliable picture of the development of depression over time, more frequent measurements are required. To limit costs, it is advisable to use a less costly mode, such as mail questionnaires, in addition to the more commonly used face-to-face interviews. However, in using both modes in one study, a mode effect occurs. Mail questionnaires yield systematically higher depression scores than face-to-face interviews in both younger old and older old males and females. It has been shown in this study that a 'sandwich formula' in which mail questionnaires are preceded and followed by a face-to-face interview can be fruitfully used in a longitudinal study, provided that a transformation is performed in order to deal with the mode effect.

Acknowledgements

This study was conducted as a part of the Longitudinal Aging Study Amsterdam (LASA), which is primarily funded by the Dutch Ministry of Health, Welfare and Sports and by the Netherlands Organization for Scientific Research (NWO), Programme Chronic Diseases – Neurological and Psychiatric Disorders.

References

Aquilino WS. Effects of Interview Mode on Measuring Depression in Younger Adults. J Off Stat 1998; 14: 15–29.

Baumgart EP, Oliver JM. Sex-ratio and gender differences in depression in an unselected adult population. J Clin Psychol 1981; 37: 570–4.

Beekman ATF, Van Limbeek J, Deeg DJH, Wouters L, Van Tilburg W. Screening for depression in the elderly in the community: using the Center for Epidemiologic Studies Depression Scale (CES-D) in the Netherlands (in Dutch). Tijdsch voor Gerontol Geriatr 1994; 25: 95–103.

Beekman ATF, Deeg DJH, Van Limbeek J, Braam AW, De Vries MZ, Van Tilburg W. Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community based sample of older adults in the Netherlands. Psychol Med 1997; 27: 231–5.

Berkman LF, Berkman CS, Kasl SV, Freeman DH, Leo L, Ostfield AM, Coroni-Huntly J, Brody JA. Depressive symptoms in relation to physical health and functioning in the elderly. Am J Epidemiol 1986; 124: 372–88.

Blazer DG. Epidemiology of late-life depression. In LJ Schneider, CF Reynolds III, BD Lebowitz, AJ Friedhoff (eds) Diagnosis and Treatment of Late Life Depression; Results of the NIH Consensus Development Conference. Washington: APA, 1994, pp 9–21.

Bradburn N. Response effects. In P Rossi, J Wright, A Anderson (eds) Handbook of Survey Research. San Diego: Academic Press, 1983, pp 289–328.

Bryson SE, Pilon DJ. Sex differences in depression and the method of administering the Beck Depression Inventory. J Clin Psychol 1984; 40: 529–34.

Cannell CF, Fowler FJ. Comparison of a self-enumerated procedure and a personal interview: A validity study. Pub Opinion Q 1963; 27: 250–64.

Deeg DJH, Westendorp de Serière M. Autonomy and well-being in the aging population I. Report from the Longitudinal Aging Study Amsterdam 1992–1993. Amsterdam: VU University Press, 1994.

Groves RM. Survey Errors and Survey Costs. New York; John Wiley & Sons, 1989.

Heeringa SG, Liu, J. Complex sample design effects and inference for mental health survey data. International Journal of Methods in Psychiatric Research 1998; 7: 56–65.

Herzog AR, Kulka RA. Telephone and mail surveys with older populations: a methodological overview. In PM Lawton and AR Herzog (eds) Special Research Methods for Gerontology. New York: Baywood Publishing Company, 1989, pp 63–89.

King DA, Buchwald AM. Sex Differences in Subclinical Depression: Administration of the Beck Depression Inventory in Public and Private Disclosure Situations. J Personal Soc Psychol 1982; 42: 963–9.

Kohout FJ, Berkman LF, Evans DA, Cornoni-Huntley J. Two

Shorter Forms of the CES-D Depression Symptoms Index. J Aging Health 1993; 5: 179–93.
Leeuw de ED. Data Quality in Mail, Telephone, and Face to Face Surveys. Thesis, University of Amsterdam, Amsterdam, 1992.
Nunnally JC. Psychometric Theory. New York: McGraw-Hill, 1978.
Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977; 1: 385–401.
Radloff LS, Teri l. Use of the CES-D with older adults. Clin Gerontol 1986; 5: 119–36.
Schwarz N, Strack F, Hippler HJ, Bishop G. The impact of administration mode on response effects in survey measurement. Appl Cogn Psychol 1991; 5: 193–212.
Smit JH, Van den Eeden P, Deeg DJH, Beekman ATF. The effects of interviewer and respondent characteristics on answer behavior in survey research: a multilevel approach (in Dutch). Sociologische Gids 1995; 42: 285–300.
Sudman S, Bradburn NM. Response effects in surveys: A review and synthesis. Chicago: Aldine, 1974.

Correspondence to: S.W. Geerlings, Longitudinal Aging Study Amsterdam, De Boelelaan 1081 C (SCW-05), 1081 HV Amsterdam, The Netherlands.
Tel. 020-4446769. Fax 020-4446775.
Email S.Geerlings @SCW.VU.NL.